VOCAL FRY AND HARSHNESS

By

JOHN FREDERICH MICHEL

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

August, 1964

ACKNOWLEDGMENTS

The author expresses his sincere appreciation to Dr. Harry Hollien for his encouragement, guidance and assistance, not only during the course of this dissertation, but throughout a rewarding and enriching academic association. The author feels extremely fortunate to have had this opportunity to study with him.

In addition, the author acknowledges with appreciation the cooperation of the University of Florida Medical Center, the Illinois Eye and Ear Infirmary, the Manhattan Veterans Administration Hospital and the National Hospital for Speech Disorders for making their facilities available to obtain subjects.

Finally, the author is especially grateful to his wife, Lorraine, for her patience, encouragement and invaluable aid throughout his course of graduate study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

REVIEW OF THE LITERATURE AND STATEMENT OF THE PROBLEM

The need for more reliable and precise descriptions of voice quality currently constitutes one of the most important problems confronting speech scientists and speech pathologists. Most authors (1) (4) (14) (30) (58) (62) agree that the normal voice should exhibit adequate loudness and pitch and a clear pleasant tone. While descriptions of loudness and pitch can be related to the physical continuums of intensity and frequency, descriptions of quality are inconsistent and tenuous due to a lack of established physical correlates. This is true not only of normal voice quality but also applies to deviant quality. As a result, attempts to describe normal and deviant voice quality resulted in the formation of many ambiguous categories (20) (22) (26) (45) (47). For example, Rush (45) categorized voice quality as natural, whisper, falsetto and orotund. Later, Goldsbury and Russell (22) added nasal or twangy, pectoral or heavy-hollow, and harsh or husky-grating to Rush's basic four. Another system was proposed by Hamill (26) who suggested pure tone, orotund, aspirate, gutteral, pectoral, nasal and falsetto as a set of voice quality categories. While Fulton and Trueblood (20) generally agreed with the classification system of Hamill, they substituted the term normal for pure tone. It is apparent that this lack of agreement in voice quality

categorization has contributed in a substantial manner to the ambiguity of voice quality descriptions. In turn, effective communication among research and clinical workers has been greatly limited.

Even with current writers (4) (14) (58) (62), consistency is not easily found. Anderson (1) suggests the terms breathiness, hoarseness and huskiness, harshness and stridency, and throatiness as labels of voice disorders. Then, however, he uses the term harsh as a partial description of huskiness and throatiness and breathiness to describe hoarseness. This would seem to indicate no clear cut or definitive boundaries among his categories. Van Riper (57) writes that voice quality defects "include excess nasality, denasality, throatiness, harshness and all the other descriptive terms which may be used to denote peculiarities of timbre." This statement emphasizes the lack of agreement and precision among current descriptions of deviant voice quality. Such confusion is stressed by Joos (32) who has suggested that the necessity of expression or description has led to the use of terms for which inadequacy is often overlooked. He stated further that a lack of sophisticated instrumentation has caused the assignment of these terms to be based on subjective impressions rather than controlled research.

The classification developed by Curtis (14) and Fairbanks (17) seems to be the most concise and least ambiguous available. They define four major classes of voice disorders: nasality, breathiness, hoarseness and harshness. However, there is still considerable misunderstanding with respect to the latter two classes. Support of this statement may be found in a study by Thurman (54) who reported that "hoarse and harsh were used to describe the same voice" more

often than any other pair of categories. He also mentioned the occurrence of "the greatest confusion between the terms 'hoarse' and other terms and between 'harsh' and other terms."

A related problem concerns one of the most common confusions of this type; the level of differentiation between clinical harshness and the normal vocal phenomenon of vocal fry. For example, Moore and von Leden (36) suggest that vocal fry and harshness describe the same voice quality while other authors (10) (27) (29) regard vocal fry as a separate entity. A complete discussion of pertinent research will be found in the following sections. However, it is relevant at this point, to indicate that the purpose of this investigation is to establish operational definitions of vocal fry and harshness in conjunction with and based on physical measures.

Research in Vocal Fry

In this investigation, vocal fry is perceptually defined as a quasi-periodic series of relatively distinct pulses. However, this voice quality has not been conclusively determined. Some authors (38) (57) have alluded to the possibility that it may be a normal function, while others (36) (58) have stated that it is an abnormal voice quality relating to harshness.

In a recent research grant proposal, Hollien (27) postulated that "vocal fry is a normal characteristic of the human voice." He also suggested that "vocal fry can best be represented as a register of fundamentals occurring in frequencies below the normal pitch register just as falsetto is a register occurring in the higher frequencies of the pitch continuum of human phonation." Support for these

contentions is evident when several factors are considered. First, vocal fry can be produced by virtually everyone, especially after practice. In fact, Moore and von Leden (36) state that it is a quality "uttered by everyone occasionally and by some people most of the time." Secondly, the frequency of vocal fry can be varied within a subject. Hollien and Wendahl (29) found that subjects were able to produce vocal fry within a large range of frequencies, most of which fell below their normal register. Finally, in a pilot study, the author found that individuals were able to produce a range of frequencies in these separate registers: normal, falsetto and vocal fry.

Irrespective of the contention that frequency may be a determinant of vocal fry, Coleman (10) reported that the damping of the wave rather than its repetition rate was the important factor in its perception. Support for Coleman's finding was reported by Wendahl, Hollien and Moore (61). In this study a large number of phonellegrams of individuals producing vocal fry were measured; it was found that a highly damped wave-form characterized this quality. Therefore, to meet the criterion of consistent perception throughout the register, some alteration must take place in the larynx to maintain the damped, pulse-like character of vocal fry at frequencies not normally perceived as pulse-like. Supportive evidence for such an alteration is supplied by the high speed motion picture research of vocal fold vibration reported by Timke, von Leden and Moore (55). They found vocal fry to be the result of a pulse-like opening and closing of the vocal folds. Further evidence that damping is the crucial parameter for the identification of vocal fry was supplied by Wendahl *et al*.

(61). They coupled an electrical laryngeal analog (LADIC) to a vowel synthesizer and generated wave-forms very similar to those of vocal fry. When this signal was compared to human vocal fry phonation, the authors comment on the similarity of the auditory event by stating: "Even the experimenters, who have had considerable experience in listening to tape recordings of the signal obtained from each source (vocal fry and the electrical analog) were unable to perceive differences between the two." They conclude that:

> It would seem then that the primary criterion that must be met for the perception of a voice signal as vocal fry is that there be nearly complete damping of the vocal tract between successive excitations. In one case, where there are two wave-fronts occurring in rapid succession followed by a long time interval before the next set of wave-fronts, the criterion has been met that the tract be allowed to decay. In the perhaps more common case of a single pulse, the criterion is also met.

This conclusion was corroborated by Coleman (10) who reported that even for frequencies around 90 cps, "there was little relationship between the repetition rate and the perception of vocal fry," but that "when the damped wave has been allowed to decay nearly to zero before the filters are restruck, vocal fry will be perceived."

Research in Harshness

"Harshness" is a subjective term applied to a perceived phenomenon. For this investigation, the definition supplied by Curtis (14) will be employed. He states that "harsh voice quality has an unpleasant, rough, rasping sound," and "if often heard in people for whom voice production seems to be a considerable effort or strain."

Systematic investigations of this phenomenon have generally taken two main approaches: judgmental and instrumental. In the judgmental studies, listeners have been asked to report their

perceptual impressions of the harsh stimulus, while instrumental studies have had as their primary goal the analysis of physical and acoustical properties of the stimulus. The literature will be reviewed with these two approaches in mind in order to provide a rationale for the procedures that will be followed in this research.

Judgmental Studies. Studying the influence of certain vowel types on the degree of perceived harshness, Sherman and Linke (51) had fifteen subjects, whose voices were diagnosed as clinically harsh, read six different passages. Each passage had a predominance of vowels from one of six categories: high, low, front, back, tense and lax. Thirty-five judges rated the passages on a seven-point scale for degree of harshness. Analysis of the data revealed statistically signigicant differences with the low vowels being judged more harsh than the high vowels. Moreover, although not statistically significant, there was a tendency for the lax vowels to be less harsh than the tense vowels, while no differences were noted between the front and back vowel groups. The reliability of these ratings had been established (Pearson r = .97) between practice and test ratings; however, no validity measures were reported. In 1955, Brubaker and Dolpheide (8) investigated the influence of consonants and vowels on the judged voice qualities of syllables and reported results very similar to those of Sherman and Linke.

In order to obtain judgments of harshness uncontaminated by irrelevant factors such as articulation, inflection, general effectiveness, etc., Sherman (49) played the Sherman and Linke (51) tapes to four separate groups with thirty to thirty-five judges in each group. All groups heard the tapes played forward and backward,

each time making judgments as to the severity of the voice quality present. Mean Q values for the two presentations did not differ significantly, indicating validity of the previous judgments.

Sherman and Jensen (50) attempted to evaluate perceived harshness for harsh and normal voices as a function of oral reading time. They found that vocal abuse or strain (defined as prolonged oral reading) did not result in judgments of increased severity of harshness. Their major conclusion was that vocal abuse, if present in harsh voice production, did not produce physiological changes in the larynx that resulted in judgments of increased harshness.

Rees (44), in 1958, instructed twelve individuals with harsh voices to record nine vowels 1) in isolation, 2) in CVC contexts with eight consonants, and 3) following /h/. Thirty-two graduate students then ranked the randomized recordings on a 1-7 scale of harshness. An analysis of the rankings indicated that the low vowels were reported as being more harsh than high vowels, vowels in voiced and fricative consonantal environments more harsh than those in voiceless and stop-plosive environments, and isolated vowels more harsh than vowels preceded by /h/. The finding that low vowels are reported as more harsh than high vowels is in agreement with the previously mentioned Sherman and Linke study (51).

It is important to note that in studies of this nature, judges were trained on material similar to that which they were to evaluate in the test situation. The effects of such training are not known. It is also difficult to know the exact nature of the signals the listeners were judging or whether the voice qualities presented were homogeneous with respect to their auditory characteristics. It

would seem, therefore, that investigations are needed in which listeners are required to make only quality discriminations without prior training, without having to define the quality and without having to assess severity. In this way, the listeners would be responding to quality similarities or differences and would not have to specify the basis of their judgment.

Further, in research of this kind, the distinction must be made between clinical harshness and simulated harshness. The former may be characterized in terms of the Curtis definition and is typified by a functional or organic problem. Simulated harshness can be defined as a voice which is perceived to be rougher than normal but exhibits no pathology. Unfortunately, simulated harshness probably consists primarily of vocal fry. The possibility that this distinction has not been made in previous investigations may contribute to the confusion between the voice disorder of harshness and the normal function of vocal fry. Moreover, that this distinction has not been made also could account for at least some of the inconclusive results reported in studies of harshness. Thus, even though these perceptual studies are valuable, it must not be assumed that they describe the physical or acoustical properties of harshness, but rather only certain aspects of the severity of harsh-like sounds. In other words, they offer inadequate bases for the precise description of harshness.

<u>Acoustical Studies</u>. One of the early acoustical studies of harshness was reported by Van Dusen (56) in 1941. His study concerned what he referred to as "metallic" or harsh voice quality. Using the Henrici analyzer, he studied the spectra of five vowels at five pitches and found that harsh female voices had most of the acoustic

energy contained in the fundamental, while normal voices of females had most of the energy in the lower overtones. Harsh male voices, however, had a spread of energy throughout the spectrum, more so than normal voices of males. These findings could indicate that 1) either the metallic or harsh voice manifests itself in different ways in males than in females, 2) the two groups did not exhibit homogeneous voice quality, or 3) these are not factors actually related to harshness. Fairbanks (17) showed by visual inspection of sonagrams, spectral differences among simulated deviant voice qualities as produced by one speaker. He reported that both harshness and hoarseness had relatively well defined first formant regions. A difference appeared, however, in the upper portions of their spectra as harshness had vertical striations through the relatively distinct formant regions indicating a pulse-like signal. The upper formants of hoarse quality were indistinct and exhibited more of a spread of energy throughout the spectrum and little if any pulse characteristics. In still another study, Thurman (54) found that he could scale various voice disorders but could not differentiate among them using the spectrographic technique of analysis. The results obtained by Van Dusen with the Henrici analyzer gives relatively precise information concerning the harmonic structure of a single wave whereas the sonagraph utilized by Fairbanks provides somewhat less precise information concerning exact harmonic structure but does so over successive waves rather than a single wave adding the dimension of time. Comparing Van Dusen's single wave to Fairbanks' succession of waves, it is noted that the male metallic voice appears to be most similar to the hoarse voice rather than the harsh voice. This similarity is not so certain,

however, knowing that Fairbanks describes hoarseness as combining the features of harshness and breathiness. Attempting to explain the differentiation achieved by Fairbanks but not by Thurman, it could be stated that Fairbanks' results were based on one speaker simulating voice disorders whereas Thurman used more than one speaker and several judges. Such varied and diverse results among studies also may be due in part to the refinement of instruments and techniques in spectrographic investigations (18) (40) (42). Although spectral analysis is not a procedure used in the present investigation, these studies were discussed in order to indicate different research methodologies utilized by different investigators.

Several studies of harshness have been completed that used oscillographic or phonellegraphic techniques. Bowler employed a custom-made galvanometric oscillograph in order to carry out a fundamental frequency analysis of harsh voice quality. He found that judgments of harsh voice quality were often associated with frequency breaks, falling inflections, or lower-than-normal fundamental frequency levels. Conversely, Duffy (16) found instances of frequency breaks on his phonellegrams but reported no consistent corresponding judgments of harsh or rough quality. These are among the very few studies which attempt to correlate judgmental and physical measures. That these findings lacked agreement emphasizes the difficulty in dealing with these concepts and points to the need for systematic, controlled research in this area.

Another relevant area of interest is frequency perturbations or wave-to-wave fluctuations in the period of the vocal signal. Discussions of speaking fundamental frequency generally have been

conducted with the assumption that the frequency at which the vocal folds vibrate is a consistent quantity; that is, the periods in any sequence of cycles are very similar if not exactly equal. However, supportive evidence for this contention is not readily available. In fact, as early as 1906, Scripture (48) studied speech curves and found acoustical data indicative of wave-to-wave variation in frequency and amplitude. Twenty years later, Simon (52) reported the variability of consecutive wave lengths in vocal and instrumental sounds. While it must be recognized that some of this variability may have resulted from measurement error, these data are more than likely accurate. In fact, Stevens and House (53) and Flanagan (19) allow for the possibility of variation in the laryngeal tone by referring to it as being "quasi-periodic." Another reason to suspect that some type of variation occurs in the vocal folds is the proposal by Lieberman (33), who states that the introduction of wave-to-wave variation into the fundamental frequency of synthesized speech possibly could enhance its quality.

The relationship between perturbations and laryngeal pathology has been demonstrated by Lieberman (34) who noted that as the disorders increasingly interfere with the normal vibratory patterns of the vocal folds, perturbations increase. Moore and Thompson (35) agree and state that from their research random fundamental frequency perturbations exist in hoarse voices. Another of Lieberman's findings was that the longer the fundamental period, the greater the perturbation. This was true for normal voices but was even more pronounced for pathological larynges characterized by growths (i.e. tumors, polyps, etc.). Although Lieberman reported no

data concerning the correlation between the extent of perturbations present and perceived voice quality, Moore and Thompson stated that increases in the amount of random variability resulted in judgments of increased severity of hoarseness Empirical data to this effect were supplied by Wendahl (60) who programmed an electrical laryngeal analog (LADIC) to produce ten conditions of random variation ±1 to ±10 cps around a median frequency. Over five hundred judges participated in a paired-comparisons procedure to determine which of the sounds were rougher. His results indicate that the "magnitude of roughness judgment was directly related to frequency differences between successive cycles." That is, variation of ±10 cps around the median frequency of 100 cps was regarded as more rough than a ±1 cps variation.

Summary

From the material presented in this chapter, it is apparent that there is a need for the investigation and correlation of judgmental and acoustical features of normal as well as of abnormal voice qualities. Moreover, it is equally apparent that much of the inconsistency that currently exists in the area of voice quality can be traced to the continued use of terms based primarily on subjective impression rather than on the results of controlled research. It is also evident that few studies have attempted to define, measure and relate at least some of the physical aspects of voice quality to the perceptual impression. Apparently the absence of such studies in vocal fry and harshness has led to the confusion concerning the use of these terms.

The recent interest in the phenomenon of vocal fry has led to some speculation as to its cause, function and characteristics, whereas the less identifiable term harshness, has been used to describe a variety of vocal qualities resulting from a variety of conditions (1) (2) (3) (4) (6) (7) (9) (15) (23) (24) (25) (31) (37) (43) (46) (59). From this multiplicity of proposed causes of harshness, it is quite apparent why subjective terms rather than operational definitions have been used to describe this quality. It would seem therefore that if accurate judgments could be made in the auditory differentiation of vocal fry and harshness, multiple instrumentation then could be employed to define more precisely their essential parameters.

Purpose

The purpose of the present investigation was two-fold. The first purpose was to build operational definitions of vocal fry and harshness. The second was to determine whether the resulting definitions were actually based on discriminable operations, both of a judgmental and acoustical nature.

Within the scope of these broad questions, several specific questions were asked.

1. Can trained and untrained listeners differentiate between vocal fry and harshness?
2. Are there fundamental frequency differences between vocal fry and harshness?
3. Are there differences between vocal fry and harsnness regarding the extent and range of frequency perturbations?
4. Are there differences between vocal fry and harshness regarding a measure of aperiodicity in the vocal signal?

The following chapter will describe the procedures by which these questions were investigated.

CHAPTER II

PROCEDURE

In order to obtain auditory judgments of vocal fry and harshness, and to investigate the acoustic parameters associated with these phenomena, the following procedures were carried out.

Subjects

Twenty male speakers were selected as subjects. Ten individuals had normal voices and also were able to produce easily a constant rate of distinct vocal fry, as determined by two experienced judges. The voices of the second ten subjects had been diagnosed as clinically harsh. All subjects were over eighteen years of age, in good general physical condition and of at least average intelligence.

Subjects for the first (normal) group were selected from the faculty and graduate students associated with the Communication Sciences Laboratory at the University of Florida and from a population of approximately forty undergraduate students. Obtaining normal and vocal fry productions from the same individuals was judged desirable in order to assure that the vocal fry samples were produced by individuals with normal rather than pathological voices.

The clinical (harsh) group was chosen from a master group of sixteen individuals judged by the investigator to exhibit clearly harsh or rough sounding voices. Recordings were made of their voices, randomized, and then played to a panel of four faculty members who have

had extensive experience with voice research and/or voice disorders. These listeners judged whether or not each voice constituted a good sample of harshness, their judgments being based on the definition of harshness as supplied by Curtis (17). For the subjects selected, it was agreed that harshness, rather than breathiness, hoarseness, etc., was the primary deviation from normal voice quality. Specifically, the panel rated each recording on a 3-point scale with 1 designating either a normal or at least a non-harsh voice, 2 representing a fair to good example of harshness, and 3 a very good example of harshness. The ten recordings receiving the highest ratings were chosen for investigation; these ten all received very high total scores. Moreover, all judges had given the same recordings the ten highest ratings although not necessarily in the same order.

It is of interest to note the difficulty encountered by the investigator in obtaining a harsh population for study. Outpatients from the ENT and Speech Clinics at the University of Florida Health Center were examined in order to identify and record subjects with harsh voices. This effort resulted in identification of only a small number of subjects. At this point, a number of calls were made to various clinics and universities throughout the United States. Approximately forty contacts were made in thirteen states, with the most frequent response being that no patients with harsh voices were available. Finally, the promise of suitable subjects warranted travel to Chicago and New York where the additional subjects were obtained only after much further screening and selection. The additional evaluation was necessary as many of the voices that were specified as harsh could not meet the criteria of rough, raspy, etc. as proposed by Curtis, and

therefore were not suitable.

Since many of the hospitals and clinics at which subjects were recorded considered case histories to be confidential information, data on age, diagnosis, etiology, etc. were not available for many subjects. However, a statement of etiology was known for several subjects and included carcinoma of the larynx and recurring vocal polyps. There was also a voice that subjectively sounded as though simultaneous vibration of the vocal and ventricular folds was occurring. Complete data of this type were not obtained since the purpose of this investigation was not to correlate etiology and vocal characteristics.

Vowel and Speech Samples

The vowel and speech samples used in this research were readings of the Rainbow passage (17) and four to six second sustained phonations of the vowel /ɑ/. The choice of the vowel /ɑ/ was based on the results obtained in a study by Rees (44) who found that the low vowels were generally more harsh than the high vowels. Since the present study was concerned with harshness, it seemed appropriate to choose a vowel that would provide the best sample of this vocal condition.

Subjects possessing normal voices made two recordings of the Rainbow passage and the sustained vowel. The first recording was performed with normal phonation and the second was in vocal fry. Although there is no apparent correlation between an individual's normal and vocal fry fundamental frequencies, it must be emphasized that the normal and vocal fry data were produced by the same individuals and hence the normal-vocal fry comparisons may not be truly independent. The subjects with the harsh voices made one recording each of the

Rainbow passage and the sustained vowel. In order for any speech sample to be acceptable, the voice quality observed during the recording had to be consistently harsh.

The Rainbow passage and sustained vowel recordings were used in the judgmental procedure, the Rainbow passage in the fundamental frequency analysis and in the aperiodicity procedure and the sustained vowel was used in the perturbation procedure.

Recording Procedure

In order to obtain satisfactory recordings, practice trials were given until each subject was able to read the passage to the satisfaction of the investigator. The criteria was consistency between the voice observed during the recording and the subject's habitual quality. Ordinarily, three to four practice trials were sufficient but additional trials were permitted if necessary. Subjects were instructed to read the passage as though addressing a group of five people.

All recording was carried out using an Ampex 601 Model 674 dual-track tape recorder coupled to an Altec 150A condenser microphone with a P518A power supply. Each speaker positioned himself approximately ten inches from, and at a ninety degree angle to, the tip of the microphone and maintained this position during the recording period. This was done to eliminate possible distortion resulting from overdriven speech sounds.

After the initial vowel recording on one channel of the recorder, a 2000 cps square wave (Hewlett-Packard Model 211AR) was recorded on the second channel to provide the reference signal (time line) used in the perturbation procedure.,

Judgmental Procedure

Purpose. The purpose of this procedure was to determine whether listeners could differentiate between vocal fry and harshness. The listeners were given no information concerning the nature of the material except that they were to make voice quality judgments of vowel sounds and speech segments. No voice quality terminology was used by the investigator at any time.

Equipment. The tapes were played on an Ampex 601 Model 674 two channel tape recorder and amplified by a Heath Model AA-161 14 watt amplifier. The output of the amplifier was coupled to an AR-1 speaker system which was placed with the listeners in an IAC Model 413-A sound treated room.

Vowel and Speech Samples. The vowel and speech material consisted of two tapes of twenty randomized items. The first tape was composed of twenty samples of the vowel /ɑ/, ten spoken by the normal group in vocal fry and ten by the harsh group. For this tape, 1 to 1.5 second samples were edited from the sustained vowel phonation recorded by subjects in the vocal fry and harsh groups. The second tape was composed of twenty speech segments from the Rainbow passage, ten spoken by the normal group in vocal fry and ten by the harsh group. Three to five second phrases, being representative of their usual degree of harshness, were edited from the Rainbow passage of each harsh subject, matching phrases being edited from each vocal fry passage. The resulting samples of vocal fry and harshness were randomized for presentation.

Listeners. Three groups of listeners were used for the judgmental portion of the investigation.

Group One consisted of ten listeners who had a great deal of experience in making auditory judgments, primarily of a psychophysical nature. Hereafter they will be referred to as the "Experienced" group.

Group Two consisted of ten individuals working for a degree in speech pathology and who had completed at least one course dealing with voice disorders. They will be referred to as the "Speech Pathology" group.

Group Three consisted of ten individuals with no experience or formal training in speech or voice disorders. They were, however, all undergraduates currently enrolled at the University of Florida and will be referred to as the "Untrained" group.

All listeners exhibited essentially normal hearing meeting the criterion of less than a 5 db hearing loss binaurally at 125, 250, 500, 1000, 2000 and 4000 cps.

Procedure. Each set of ten listeners entered the IAC room and performed the judging tasks as a group. They were asked not to consult with each other at any time before or during the test procedure. Two specially prepared answer sheets (see Appendix A), one for each task, were given to each listener.

Briefly, the listeners were instructed to listen to the twenty voice samples and divide them into two groups of ten each. The basis of this division was to be the quality of the voice; they were to place ten voices of similar "quality" in one group and ten voices of the other "quality" in the other group. The first tape

(vowels) was played once to familiarize the listeners with the nature of the material and to allow them time to establish some basis for their judgments. The tape was played the second time and the listeners recorded their judgments on the answer sheets. The tape was then played a third time permitting the listeners to confirm their responses. The same procedure was followed with the tape containing the speech samples The obtained data were then analyzed by chi-square in order to evaluate how consistently the two types of voices were differentiated The expected frequency was derived from that which would be anticipated if all listeners were able to discriminate vocal fry and harshness.

Fundamental Frequency Procedure

Purpose. The purpose of this procedure was to test the hypothesis that no differences exist between the fundamental frequency levels of vocal fry and those of harshness

Equipment. The Fundamental Frequency Indicator (FFI) and the phonellegraph were used to obtain fundamental frequency measures. FFI is an automatic device designed primarily to extract the fundamental period from complex waves. Briefly, it consists of one-half octave low pass filters and high speed switching circuits. The fundamental frequency information extracted is then stored on magnetic tape as a continuously varying sine wave. The speed of the tape is then reduced by a factor of four and the output fed into an electronic counter which measures the period of the first cycle occurring every 33 milliseconds. A teletype prints these data on paper tape, from which IBM cards are punched and the data they contain analyzed

by the University of Florida's IBM 709 computing system. The program includes the conversion of each period score to a frequency score and the computation of the geometric mean frequency level and standard deviation in semitones.

The Hollien-Malcik (28) modification of the phonellegraph was also used for this procedure. This device provides a visual trace on photo-sensitive paper from which measures of fundamental period can be obtained and converted to frequency.

<u>Material</u>. The Rainbow passages produced by the subjects with normal voices were analyzed by FFI. The passages spoken in vocal fry by the normal subjects and those produced by the harsh subjects were analyzed by the phonellegraphic procedure.

Since two analysis procedures were used their reliability and validity had to be assessed. Such evaluation had been carried out previously on normals and was satisfactory. In order to determine the accuracy with which FFI could track fundamental frequency in vocal fry and harshness, such data were obtained also for these subjects. The Rainbow passages of the vocal fry and harsh subjects were analyzed both by the phonellegraph and FFI procedures. Correlations were then calculated between the results obtained by the two procedures. It was found that the vocal fry data from FFI were not usable due to the fact that the lower cut off limit of 50 cps was exceeded by the vocal fry signal. However, when the harsh samples were analyzed by both procedures (FFI and the phonellegraph), a correlation of .99 was observed. Since the primary purpose of this investigation was to study vocal fry and harshness, the phonellegraphic procedure, common to both vocal fry and harshness was used.

Procedure and Measurement. The normal samples of the Rainbow passage were processed by FFI and the 709 computer system. Thus, values were obtained for the normal fundamental frequency and standard deviation, however, the program did not include the computation of the total or 90% ranges.

The oral reading passages of the vocal fry and harsh subjects were re-recorded onto discs using a Presto 6N disc recorder and phonellegrams made using the phonellegraph described above. The phonellegrams were divided into twenty intervals of equal length and measurements made of the average period within each Interval. These period values were arranged into a frequency distribution from which measures of the mean fundamental period were made and converted to frequency. Measures were also made of the standard deviation and total and 90% ranges. The data from both procedures were then combined and an analysis of variance was computed to analyze differences among normal, vocal fry and harshness with respect to the measures of fundamental frequency.

Perturbation Procedure

Purpose. The purpose of this procedure was to test the hypothesis that there are no differences between vocal fry and harshness in the mean and range of fundamental frequency perturbations.

Equipment. The tape recorded vowels were replayed on an Ampex 601 Model 674 two channel tape recorder and fed into a Tectronix Type 502 dual-beam oscilloscope. The oscilloscopic view system of a Fastax high-speed motion picture camera was used to allow the oscilloscopic traces to be photographed.

Material. The material used was the tape recordings of the sustained vowel produces by the harsh subjects and the normal subjects phonating in vocal fry.

Procedure. The sustained vowel /ɑ/ was recorded on one channel of a two channel recorder after which a 2000 cps square wave was recorded on the other channel. These signals were fed simultaneously into separate channels of a dual-beam oscilloscope. The time base amplifier was disconnected from the circuit so that only the vertical deflection amplifier had an effect on the inputs to either channel. Following this, the right angle oscilloscope lens attachment from a Fastax camera was focused on the face of the scope. With the main lens capped to eliminate framing, photographs were made with the oscilloscopic lens opening at f/2 and a film-to-scope distance of twenty-eight inches. The tape was started and after allowing two to three seconds for the recorder to stabilize, the camera was fired at a speed of 1000 frames per second, exposing a 100 foot roll of Tri-X black and white film in approximately four seconds.

In order to make measurements, a film viewing box was constructed so that approximately sixteen inches of film (fifty-four frames) could be viewed at one time. When this length of film was illuminated, four to seven consecutive cycles were visible — depending on the frequency of the signal. To enhance accuracy of measurements it was necessary to indicate the termination of one cycle and the beginning of the succeeding one. Due to the speed of the film, the individual cycles of some of the films were lengthened to the extent that "leading edges," as described by Lieberman (33) (34), of amplitude peaks were not clearly discernible and yet the

repetitious character of the signal was maintained. To reduce error in the definition of cycle boundaries, a line was scribed on the emulsion side of the film parallel to the edge and intersecting the ascending and descending slopes of the largest peak in the wave. The distance from slope to slope was measured along this line and divided in half. From this point, a perpendicular was drawn to intersect the 2000 cps reference signal. Fifty consecutive cycles or a one second length of phonation were so marked and measurements made of the number of 2000 cps cycles occurring for each cycle of the speech signal. In other of the films, distinctive features were discernable and used as points from which to draw the perpendiculars. The period of the time-line was checked occasionally and found to be a consistent length of 12/64ths of an inch. Thus when the perpendicular line did not intersect the spiked, leading edge of the time line, the number of 64ths of an inch was converted to hundredths of an inch and added to the number of complete time line cycles contained within the length of the vocal cycle. The frequency and range in cps were then computed for each of the fifty consecutive speech cycles. Finally, to obtain a measure of the mean extent of cycle to cycle variation, the differences between adjacent cycles were averaged for each subject and comparisons made between groups. Analyses of variance were computed among the mean perturbations and among the range or envelope of perturbations.

Aperiodicity Procedure

Purpose. The purpose of this procedure was to test the hypothesis that no difference exists between vocal fry and harshness

regarding the percent of aperiodicity or unmeasureable phonation in the vocal signal during speech.

Material. The phonellegrams of the vocal fry and harsh Rainbow passages provided visual tracings from which the amounts of measurable and unmeasureable phonation could be determined. In addition, phonellegrams were made and measured for the passages recorded in normal voice quality.

Procedure. The total phonation time of each sample was measured omitting all silent portions. Following this, the aperiodic or unmeasurable portion of this total was then determined and the following ratio formed: aperiodicity/total phonation. In this way, it was possible to determine the percent of total time in which measurable phonation was not present. A test for the difference between two uncorrelated percentages was computed.

CHAPTER III

RESULTS

This chapter presents the results obtained by the judgmental and acoustical procedures described in Chapter II. Raw data and the results of statistical analyses are included.

Judgmental Results

As previously described, the listeners were asked to listen to two tapes, one composed of twenty vowel segments and the other composed of twenty speech segments. As stated, the listeners were given no information other than directions that for each tape their task was to divide the twenty items into two groups of ten using quality alone as the basis for their division. Thus each completed answer sheet was scored with 10 "A" responses and 10 "B" responses, presumably with all the "A" items being similar in quality and all the "B" items being similar in quality. If a listener correctly placed all the vocal fry samples and all the harsh samples in their respective groups, the score for that listener would be 10/10. If, however, an "A" item was included in the "B" group, this forced a "B" item to be included in the "A" group making the score 9/9.

Using this procedure for both tapes, the scores for the three groups of 10 listeners each (Experienced, Speech Pathology and Untrained) were tallied. Table I presents the results of the quality discrimination judgments by the three groups. Differences between

TABLE I.--Scores and X^2 values for each of the listener groups. The No. Correct column indicates the number of correct responses for the harsh/vocal fry discrimination.

Listeners	N	No. Correct (vowels)	X^2*	No. Correct (speech)	X^2*
Experienced	10	95/95	.5	100/100	.0
Speech Path.	10	97/97	.3	100/100	.0
Untrained	10	93/93	.2	100/100	.0

$*X^2_{.01} = 21.67$, df = 9

the observed responses by the listeners and those expected for the vowel tape were of small magnitude, chi-square values being .5, .3, and .2 for the Experienced, Speech Pathology and Untrained listeners respectively. No errors were observed for the speech tape resulting in all chi-square values of zero. For significance at the .01 level, a value of 21.67 was required. Thus it would appear that listeners, regardless of their degree of sophistication or training, can discriminate between vocal fry and harshness using differences in quality as the basis for this discrimination. It was interesting to note a very slight trend in the number of errors made on the vowel tape. The Untrained listeners made the fewest errors, followed by the Speech Pathologists and finally the Experienced listeners who made the most errors. Due however, to the high degree of accuracy shown by all groups, the shallow slope of the trend was not felt to be of import. That no errors were observed for the speech segments strongly indicates that vocal fry and harshness can be discriminated in running speech. This finding would seem to be of substantial significance in the diagnosis of voice disorders.

Fundamental Frequency Results

Table II presents the results of the fundamental frequency measures of central tendency and variability for each individual subject and for each group. The data on the normal subjects were obtained from the Fundamental Frequency Indicator (FFI) while the data for the vocal fry and harsh subjects were obtained by the phonellegraphic technique.

The mean fundamental frequency observed for vocal fry subjects was 36.4 cps with a range of 30.7 to 43.7 cps. These figures were

TABLE II.--Measures of central tendency and variability for normal, vocal fry and harsh subjects. The normal subjects were analyzed by FFI and only the mean and standard deviation were obtained. The phonellegraph was used to analyze the vocal fry and harsh subjects and all measures were obtained.

Quality		Mean Fundamental Frequency in cps	Standard Deviation in semitones	Total Range in semitones	90% Range in semitones
A. Normal Phonation					
1.	N-1	100.3	2.7		
2.	N-2	122.4	3.1		
3.	N-3	103.4	2.7		
4.	N-4	122.0	4.0		
5.	N-5	119.7	3.8		
6.	N-6	100.7	3.1		
7.	N-7	115.4	3.6		
8.	N-8	125.1	2.8		
9.	N-9	102.3	3.3		
10.	N-10	94.8	2.2		
	Mean	110.6	3.1		
B. Vocal Fry Phonation					
1.	N-1	43.7	3.3	26.2	11.9
2.	N-2	36.1	3.9	22.6	11.3
3.	N-3	35.1	5.0	26.1	15.9
4.	N-4	30.7	4.9	25.7	14.7
5.	N-5	33.9	3.1	23.6	9.7
6.	N-6	40.8	4.7	26.5	14.6
7.	N-7	33.7	4.8	25.5	14.2
8.	N-8	38.7	5.3	29.5	16.1
9.	N-9	34.1	3.9	27.6	11.8
10.	N-10	30.9	5.3	23.1	15.0
	Mean	36.4	4.4	25.6	13.5
C. Harsh Phonation					
1.	H-1	103.7	3.4	30.1	11.4
2.	H-2	105.3	2.5	14.8	7.9
3.	H-3	153.4	4.9	29.8	17.4
4.	H-4	107.5	3.1	23.3	9.7
5.	H-5	106.4	4.4	35.4	14.4
6.	H-6	106.0	3.3	26.8	10.4
7.	H-7	113.6	2.2	18.6	6.7
8.	H-8	125.8	*	31.6	18.3
9.	H-9	180.0	3.4	24.9	12.8
10.	H-10	118.8	2.8	21.0	8.5
	Mean	122.1	3.3	25.6	11.8

*Due to a bimodal distribution, this measure could not be used.

compared with a mean fundamental frequency of 122.1 cps (range: 103.7 to 180.0 cps) for the harsh voices. The normal voices had a mean fundamental frequency of 110.6 cps. The data would appear to demonstrate that the mean fundamental frequency of vocal fry is consistently lower than that of harsh and normal voices while the difference between the harsh and normal voices is small.

Table III presents the results of an F ratio computed to determine the statistical significance among these mean fundamental frequencies. A value of 82.39 indicated a high level of significance among the three means. The results of the calculation of the separate t values are contained in Table IV. Values of 6.47 between vocal fry and harshness and 19.32 between vocal fry and normal demonstrate the presence of highly significant differences in fundamental frequency. A value of .79 between normal and harsh quality indicates no difference in fundamental frequency. In order to avoid Type I errors, it was felt that a level of .01 should be met or exceeded before significance was assumed. In any event, the fundamental frequencies observed in vocal fry are significantly lower than those of either normal or harsh voice quality, indicating a parameter along which vocal fry and harshness can be differentiated. No such difference was found between harsh and normal voices, indicating that they must be differentiated by other means.

Measures of variability revealed standard deviations of 4.4, 3.3 and 3.1 semitones respectively for vocal fry, harsh and normal voices. Total range measurements were identical at 25.6 semitones for vocal fry and harshness, while 90% ranges of 13.5 and 11.8 semitones for vocal fry and harshness did not show a significant difference.

TABLE III.--Summary of analysis of variance evaluating differences among the fundamental frequencies of normal, vocal fry and harsh subjects.

Source	df	ms	F	$F_{.01}$
Between Groups	2	21,642.43	82.39	5.49
Within Groups	27	262.69		
Total	29			

F ratio: ms_{bet}/ms_{within}

TABLE IV.--Values of t for the evaluation of mean fundamental frequency for normal, vocal fry and harsh subjects.

Comparison	$\overline{X}_1 - \overline{X}_2$	df	t	$t_{.01}$
Harsh/Vocal Fry	82.1	18	6.47	2.88
Harsh/Normal	7.9	18	.60	
Normal/Vocal Fry	86.2	18	19.32	

No range measurements were obtained for the normal voices as they were analyzed by FFI which did not have those measures included in the program. It was not felt that this was a significant omission. This may be emphasized by noting the striking similarities in the range data of vocal fry and harshness in spite of differences in fundamental frequency.

In summary, it was found that vocal fry has a significantly lower fundamental frequency than either normal or harsh voices. The range data of vocal fry and harshness are quite similar.

Perturbation Results

The results of the perturbation procedure are contained in Table V. Mean fundamental frequencies for the sustained /ɑ/ phonations are 112.6, 30.1 and 102.0 cps for normal, vocal fry and harsh subjects respectively. Of interest is the rather close agreement between speaking fundamental frequency level and the same measure for the sustained vowel produced by the normal and vocal fry subjects. On the other hand, a drop of more than 20 cps was noted between speaking and sustained phonation for the harsh subjects. The reason for this is not apparent.

The mean perturbation for the normal phonation was .60 cps within a mean range of 3.8 cps. Mean perturbation for the vocal fry subjects was 2.30 cps within a range of 12.9 cps while a 1.58 cps mean perturbation and a 7.7 cps range was observed for the harsh subjects. These data show vocal fry to have greater mean perturbations and range of perturbations than harsh phonation. In like manner, harshness exceeds normal voice in both these measures.

Table VI presents the results of an analysis of variance

TABLE V.--Mean fundamental frequency, perturbation factor around the mean fundamental, limits and range of perturbations for a sustained /ɑ/ phonation. Values are expressed in cps.

Quality		Mean Fundamental Frequency of a Sustained Vowel	Mean Perturbation	Limits	Range
A.	Normal Phonation				
1.	N-1	139.9	.86	138.9 - 141.8	2.9
2.	N-2	112.0	.66	110.8 - 114.3	3.5
3.	N-3	113.8	.73	110.5 - 115.6	5.1
4.	N-4	118.6	.54	116.1 - 120.6	4.5
5.	N-5	94.4	.65	91.8 - 98.6	6.8
6.	N-6	102.8	.45	101.3 - 104.6	3.3
7.	N-7	142.2	.63	140.4 - 142.9	2.5
8.	N-8	112.9	.67	111.7 - 114.3	2.6
9.	N-9	99.0	.43	97.0 - 100.0	3.0
10.	N-10	89.9	.39	88.4 - 92.0	3.6
	Mean	112.6	.60		3.8
B.	Vocal Fry Phonation				
1.	N-1	47.2	.80	44.5 - 52.5	8.0
2.	N-2	26.2	1.38	20.5 - 29.9	9.4
3.	N-3	24.6	2.59	17.6 - 31.2	13.6
4.	N-4	19.0	1.93	13.2 - 27.3	14.1
5.	N-5	41.7	2.33	35.0 - 48.2	13.2
6.	N-6	23.8	2.33	14.0 - 29.3	15.3
7.	N-7	39.5	1.29	35.3 - 43.7	8.4
8.	N-8	27.9	2.74	22.2 - 35.0	12.8
9.	N-9	27.5	3.04	21.7 - 34.2	12.5
10.	N-10	23.1	4.58	12.8 - 34.2	21.4
	Mean	30.1	2.30		12.9
C.	Harsh Phonation				
1.	H-1	107.7	4.04	103.1 - 113.0	9.9
2.	H-2	122.0	.54	120.3 - 123.5	3.2
3.	H-3	133.7	4.56	127.2 - 140.1	12.9
4.	H-4	104.1	.67	102.1 - 107.7	5.6
5.	H-5	64.1	.57	60.4 - 65.6	5.2
6.	H-6	104.2	2.24	98.0 - 108.3	10.3
7.	H-7	135.7	1.08	132.2 - 141.6	9.4
8.	H-8	94.4	.66	92.6 - 99.5	6.9
9.	H-9	85.6	.53	83.0 - 87.6	4.6
10.	H-10	68.3	.91	61.6 - 71.0	9.4
	Mean	120.0	1.58		7.7

TABLE VI.--Summary of analysis of variance evaluating differences among the mean perturbations for normal, vocal fry and harsh subjects.

Source	df	ms	F	$F_{.01}$
Between Groups	2	7.28	6.28	5.46
Within Groups	27	1.16		
Total	29			

F ratio: ms_{bet}/ms_{within}

TABLE VII.--Values of *t* for the evaluation of mean perturbation factors for normal, vocal fry and harsh subjects.

Comparison	$\bar{X}_1 - \bar{X}_2$	df	t	$t_{.01}$
Harsh/Vocal Fry	.72	18	2.02	2.88
Harsh/Normal	.98	18	1.51	
Normal/Vocal Fry	1 70	18	3.09	

computed to test for differences among the mean perturbation values for normal, vocal fry and harshness. An obtained F of 6.28 warranted the computation of individual t scores. The major contributor to the significance of the F ratio was the normal-vocal fry difference although the vocal fry-harsh difference did approach significance with a t value of 2.08, as shown in Table VII.

Table VIII shows that an analysis of variance computed among the mean ranges of perturbations resulted in significance at the .01 level with a value of 21.88 (this actually exceeded the .001 level of 9.02). Table IX shows that the t tests among these ranges were significant at the .01 level with the exception of the harsh/normal comparison which reached significance at the .05 level.

The perturbation procedure was carried out to investigate the possibility that differences exist in the wave-to-wave variability of normal, vocal fry and harsh voice qualities. It has been shown that this is the case when all three qualities are considered, but does not support the thesis that a perturbation factor can be used to differentiate vocal fry and harshness.

In summary, it is noted that normal, vocal fry and harsh voice qualities can be differentiated when considering the mean and range of perturbations. Normal voices phonating a sustained vowel have small perturbations within a small range while vocal fry has the largest perturbation factor within the largest range. Mean perturbation and range values for the harsh voices are about midway between the normal and vocal fry values.

In reporting these data, possible sources of error necessitate mention. First, the drift of the oscillator supplying the 2000 cps

TABLE VIII.--Summary of analysis of variance evaluating differences among the range of perturbations for normal, vocal fry and harsh subjects.

Source	df	ms	F	$F_{.01}$
Between Groups	2	202.21	21.88	5.49
Within Groups	27	9.24		
Total	29	211.45		

F ratio: ms_{bet}/ms_{within}

TABLE IX.--Values of t for the evaluation of perturbation range for normal, vocal fry and harsh subjects.

Comparison	$\bar{X}_1 - \bar{X}_2$	df	t	$t_{.01}$
Harsh/Vocal Fry	5.2	18	3.31	2.88
Harsh/Normal	3.9	18	2.53	
Normal/Vocal Fry	9.1	18	5.46	

square wave reference signal may be great enough to introduce measurable variation into the results. Also, the "wow" in the tape recorder could be responsible for this similar type error when the speech and reference signals are not recorded simultaneously. Finally, measurement error is always possible and should be defined by reliability checks. To evaluate the measurement error in this investigation, independent measures were made of identical films by the investigator and an associate. A high correlation (Pearson $r = .96$) indicated a minimum of error in this procedure.

<u>Aperiodicity Results</u>

Aperiodicity/phonation time values are found in Table X. The values reported in this Table were obtained by measuring the total time in which phonation was present (omitting periods of silence) and dividing this figure into the amount of aperiodicity measured in the total time in which phonation was present. Aperiodicity is here defined as a lack of recognizable repeating wave-forms. Normal voice quality was aperiodic over 2% of the time and harsh voice quality was aperiodic over 17% of the time. No values appear for vocal fry since no unmeasurable phonation could be found. Although wave-to-wave measurements revealed period variations greater than those observed for either normal or harsh phonation, the characteristic wave-form was always recognizable. This was not true for normal and harsh voices in which definite instances of noise were observed.

A $\underline{t}$ test (as found in Garrett (24) p. 235) for the difference between the obtained percentages of 2.5 and 17.5 was not significant at the .05 level. This is probably due to the large degree of

TABLE X.--Values obtained for an aperiodicity/phonation time ratio. Phonellegraphic tracings were measured and the amount of unmeasurable phonation was divided by the total time phonation was present.

Quality		Aperiodicity/Phonation Time
Normal		
1. N-1		.0239
2. N-2		.0257
3. N-3		.0153
4. N-4		.0246
5. N-5		.0120
6. N-6		.0264
7. N-7		.0270
8. N-8		.0348
9. N-9		.0000
10. N-10		.0156
	Mean	.0250
	Range	.0000 - .0348
Harsh		
1. H-1		.2014
2. H-2		.0518
3 H-3		.4432
4 H-4		.0898
5. H-5		.1590
6. H-6		.1869
7. H-7		.1061
8 H-8		.2035
9. H-9		.1669
10. H-10		.1368
	Mean	.1745
	Range	.0518 - .4432

variation in the harsh group.

The results of this procedure lend support to the idea that vocal fry and harshness can be differentiated by the computation of an aperiodicity/phonation time ratio. Assuming vocal fry or harshness to be the only alternatives in the identification of a particular voice, the presence of unmeasurable phonation would be evidence of a harsh voice whereas the absence of unmeasurable phonation would indicate vocal fry. Unmeasurable phonation was found easily in harsh quality and somewhat more difficult to find but present in normal phonation. Of course, some unmeasurable signals were expected in all types of phonation due to the presence of fricative or noisy consonantal sounds in speech. The consistent occurrence of more unmeasurable phonation in harshness than in normal phonation or in vocal fry strongly suggests the existence of other noisy elements, possibly an extremely aperiodic laryngeal signal. That no instances of noise were found in vocal fry phonation cannot be explained and certainly warrants further investigation. It may be postulated, however, that in spite of large wave-to-wave differences in period length, the pulse-like character of vocal fry made the distance from one pulse to the next easily discernible and perhaps obscured the aperiodic trace of the consonantal sounds. In any event, the aperiodicity found in harshness tends to differentiate it both from vocal fry and normal phonation.

CHAPTER IV

DISCUSSION

Judgmental Discussion

Due to the lack of agreement concerning the relationship between the auditory impression of voice qualities and the terminology used to describe these qualities, the results of the judgmental procedure in this investigation were of special interest. An important finding was that listeners, regardless of their degree of sophistication and training, were able to discriminate between samples of vocal fry and harshness. Not only were their voice quality judgments of the vowel samples in high agreement, but 100% agreement was observed when judgments were made of the speech samples. That they were able to make these judgments to such a high degree of accuracy indicates the existence of factors unique to each of the two voice qualities. Therefore, it would seem that the identification of these factors would alleviate at least some of the confusion concerning vocal fry and harshness.

Moore and von Leden (36), from a rather thorough review of the literature concerning vocal fry and harshness, indicate that these two terms are often used interchangeably, leaving the reader with the impression that they are the same phenomenon. An example of this tendency can be found in a later article by Timke, von Leden and Moore (55) in which they describe vocal fry as a low harsh sound.

Moreover, Van Riper and Irwin (58) report that vocal fry is a distinct part of harshness and can be easily recognized. From this discussion it is not known whether a voice is considered to be harsh due solely to the presence of vocal fry or whether some other factor such as noise in the vocal spectrum also contributes to the impression of harshness.

Culver (12) states that "a noise commonly consists of a group of non-periodic pulses arising from the irregular vibration of a body or group of bodies." It is assumed that many such signals would be judged harsh-like since Culver states that they "usually produce an unpleasant auditory sensation." Several current authors (33) (34) (60) have found that the aperiodic signal does not have to be pulse-like to create the impression of harshness. It is therefore postulated that the report of a rough sound can occur under at least two conditions. The first is when a sound is aperiodic to the point of losing tonality, thus becoming noise, and the second occurs when successive waves are sufficiently damped as to allow for the perception of the individual cycles or pulses. If clinical harshness can be said to be characterized by the first condition and vocal fry by the second, it would seem that listeners would be able to discriminate between the two with a high degree of agreement. The results of this study confirm this statement.

Fundamental Frequency Discussion

The results of the present investigation offer strong supportive evidence for the notion that vocal fry and harshness can be differentiated with respect to the parameter of fundamental frequency.

In order to adequately discuss the results obtained by this procedure, several characteristics of the populations under study must be considered. The first is age. Because of the confidential nature of hospital and clinic records, the ages of the harsh voiced subjects only can be estimated. However, this group could be roughly classed as middle aged; they had an approximate age range of from 18 to 60 years. Thus, it is apparent that the normal subjects obtained from a college population and the harsh voiced subjects were not matched with respect to age. Therefore, due to the possible effects of age on the fundamental frequency of the speaking voice, comparison between these two groups is somewhat tenuous and in order to avoid the possible effects of age differences, several studies of speaking fundamental frequency (13) (28) (39) (41) were considered for the comparisons below. The studies of Hollien, Malcik and Hollien (28) and Mysak (39) were thought most appropriate for two reasons. First, they are among the most recent investigations and therefore benefited from advances in the techniques for measuring fundamental frequency. Secondly, the age groups studied in these two investigations roughly were comparable to those in the present study. Hence the vocal fry group can be compared with the results of the Hollien et al. study while the harsh subjects can be compared to the middle age group studied by Mysak. There was, however, one criterion placed on the normal subjects used in this study that was not placed on the subjects studied by others. This criterion was that the subjects had to be able to produce consistent, sustained vocal fry phonation. The relationship between this ability and normal fundamental frequency level is not known at this time.

The group of 18 year olds studied by Hollien, et al. revealed a mean fundamental frequency of 115.9 cps while the normal subjects in the present investigation had a mean fundamental of 110.6 cps. The middle aged group studied by Mysak was found to have a fundamental of 113.2 cps as compared to 122.1 cps for the harsh subjects in the current study. From comparisons among these means, it can be seen that they are all very similar. In fact, the similarities among these values are underscored when they are contrasted to the mean vocal fry frequency of 36.4 cps. The very low fundamental frequency exhibited in vocal fry further emphasized one dissimilarity between vocal fry and other types of phonation.

The fundamental frequency levels exhibited by the harsh subjects also can be compared to those reported by Bowler (5) who stated that harsh voices are characterized by lower than average "pitch" levels. This was not a finding in this investigation. To the contrary, the mean fundamental frequency of the harsh voices is 26.7 cps higher than that reported for Bowler's subjects. Even the mean fundamental frequency level for the individual with the lowest voice was over 8 cps higher than the mean fundamental reported by Bowler. A possible explanation for this may be that the subjects studied here were clinically harsh subjects and not individuals whose voices "contained examples of harshness." In other words, the voices of the subjects in this study were analyzed for their general, consistent harsh quality, while Bowler studied "examples of harshness" in voices otherwise termed normal. Moreover, the non-harsh speech of Bowler's experimental subjects revealed a median fundamental frequency of 127.1 cps. It would appear that, if the harsh and non-harsh portions of the

speech studied by Bowler were averaged, a mean frequency level would be found that more closely resembles those of the harsh voices in this study.

Support for the statement that frequency breaks one octave or more in extent accompany judgments of harshness was not found in this study. In fact, it was in only one subject that frequency breaks were observed at all and the distribution of fundamental frequencies for this particular subject was distinctly bimodal. Thus, it would seem that the frequency breaks exhibited by this individual were related to shifts between the two frequency modes that he used in spoken language. No other subject was found to have a bimodal distribution or frequency breaks an octave in extent.

Bowler also associated harshness with falling inflections at the end of sentences. Herein may be one of the more common instances that produces confusion between vocal fry and harshness. It is probable that the harshness Bowler reports as occurring with falling inflections was actually vocal fry and not harshness at all. Moreover, that this distinction usually is not made is evidenced in the statement by Duffy (16) that "If this 'rough' quality occurs in normals then harshness is not a 'quality' deviation but a 'quantity' deviation of normal usage." Thus, it should be stressed that the two terms, vocal fry and harshness often may be used synonymously. This is especially true when writers attempt to define harshness while unaware of the existence of the vocal fry register. In summary, the results of this research show that fundamental frequency is one parameter by which vocal fry and harshness can be identified, as mean fundamental frequency of vocal fry is significantly lower than that of harshness.

Perturbation Discussion

Significant differences were found to exist among the mean perturbations in sustained vocalizations of normal, vocal fry and harsh voices. The implications of these differences are not totally clear, although current investigations (34) (35) have indicated a relationship between the extent and number of perturbations and the severity of a voice pathology. From this it would seem to follow that changes in voice quality and changes in perturbation level are associated in some way. To some extent, the results of the present study support this notion as the harsh voices did show a tendency toward greater perturbations than did the normal voices.

Although a greater perturbation factor was observed in vocal fry than in normal or harsh voices, the pulse-like character of vocal fry was maintained. It is therefore postulated that even though this perturbation, or irregularity of adjacent period lengths, is associated with vocal fry, probably it is not an important factor for the recognition and identification of vocal fry. Thus, it is proposed that vocal fry may be either a regular or slightly irregular train of pulses and still be judged as vocal fry. In the harsh voice however, there is no over-riding pulse-like signal to which the ear can respond. Accordingly, it is proposed that this leaves the perturbation as one of the predominate percepts.

In summary, the findings of this procedure are that normal voices have the smallest degree of random variation around a mean fundamental frequency, harsh voices have a larger amount followed by vocal fry which has the greatest amount of random variation around a mean fundamental. Thus, it can be concluded that measures of frequency

perturbation may be used to differentiate between vocal fry and normal phonation and possibly between vocal fry and harshness. However, no such differentiation is possible between normal voice quality and harshness.

Aperiodicity Discussion

The results of the aperiodicity/phonation procedure revealed that normal phonation contains approximately 2% unmeasurable or aperiodic phonation while harsh phonation is unmeasurable or aperiodic 17% of the time. No unmeasurable phonation at all was found in vocal fry. These results strongly support the statement by Fairbanks (17) that the harsh voice is characterized by noise added to the vocal spectrum. Moreover, they support the reference to a noisy vocal tone in the Avery, et al. (2) description of a harsh voice as well as Bowler's (5) statement that harsh voice quality if often associated with instances of aperiodicity. Thus, it would appear that aperiodicity is one of the distinguishing features of harsh voice quality. In vocal fry, however, peak to peak measurements were easily made with no portions being unmeasurable. This finding emphasizes the uniqueness of the vocal fry wave-form, the closer examination of which may give some explanation as to why no unmeasurable phonation was found. In normal and harsh phonation, the introduction of noise into the spectrum caused the periodic pattern of adjacent waves to become less recognizable. In vocal fry, it is proposed that the nature of the pulse-like signal is such that it is not significantly influenced by the introduction of noise in the amounts usually found in speech.

Based on the results of this investigation, it is concluded

that harsh voice quality has considerably more spectral noise present than does either normal or vocal fry phonation.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this investigation was to operationally define, and differentiate between, the voice qualities of vocal fry and harshness. To carry out this purpose, a judgmental procedure and three acoustical procedures — analysis of fundamental frequency, perturbation and aperiodicity — were employed. Two groups of subjects were studied: Group A was composed of ten males with normal voices who easily could produce a constant repetition rate of vocal fry, and Group B was composed of ten males judged to have harsh voices. Material for analysis was recordings of a standard oral reading passage and the sustained vowel /ɑ/. The subjects in Group A recorded the passage and the vowel both in their normal voice and in vocal fry. Group B recorded the same material in a voice representative of their usual degree of harshness.

Short segments were edited from each spoken passage and randomized into a tape of twenty items, ten by the normal subjects speaking in vocal fry and ten by the harsh subjects. A tape containing the vowels was prepared in the same manner. These tapes were played to three separate sets of ten individuals composed of experienced listeners, speech pathologists and untrained listeners. These listeners divided the samples on each tape into two groups using differences in quality as their sole basis of judgment.

A fundamental frequency analysis was carried out on the reading passages for the normal, vocal fry and harsh readings of the standard passage. The Fundamental Frequency Indicator and phonellegraph were utilized. From the data obtained, measures of central tendency and variability were computed.

Two procedures were completed to identify the extent of variability within the vocal signal. Frequency perturbations of the sustained vowel were obtained by measuring photographic tracings provided by a high-speed camera and oscilloscopic system. The percent of aperiodicity or unmeasurable phonation was determined from the phonellegraphic tracings.

The major conclusion provided by this research is that vocal fry and harshness are separate entities that can be differentiated on the basis of judgmental and acoustical procedures. Moreover, the following specific conclusions were reached.

First, the essential factors relating to vocal fry are a very low fundamental frequency level, large perturbations around a mean fundamental frequency level, the absence of unmeasurable phonation and the damped wave-form reported by Coleman (10).

Finally, harsh voices are characterized by aperiodicity or noise in the spectrum, a normal fundamental frequency level and larger than normal perturbations about the mean fundamental frequency.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Anderson, V., Training the Speaking Voice, Oxford University Press, New York, 1950, 68.

2. Avery, E., Dorsey, J. and Sickels, V., First Principles of Speech Training, Appleton-Century-Crofts, Inc., New York, 1928, 56.

3. Barnes, H. and Smith, L., Speech Fundamentals, Prentice-Hall, Inc., New York, 1953, 116.

4. Berry, M. and Eisenson, J., The Defective in Speech, Appleton-Century-Crofts, Inc., New York, 1942, 210-213.

5. Bowler, N., A Fundamental Frequency Analysis of Harsh Vocal Quality, Speech Monographs, 31, 1964, 129-134.

6. Brodnitz, F., Voice Problems of the Actor and Singer, Journal of Speech and Hearing Disorders, 19, 1954, 322-326.

7. Brown, H., Standardization of Vocal Training from the Teacher's Standpoint, Laryngoscope, 23, 1913, 21-28.

8. Brubaker, R. and Dolpheide, W., Consonant and Vowel Influence upon the Judged Voice Quality of Syllables, Journal of the Acoustical Society of America, 27, 1955, 1000.

9. Cadawallader, B., Elimination of Speech and Voice Defects, Laryngoscope, 24, 1914, 986-992.

10. Coleman, R., Decay Characteristics of Vocal Fry, Folia Phoniatrica, 15, 1963, 256-263.

11. Cornut, G. and Lafon, J., Vibrations Neuromusculaires des Cordes Vocales et Theories de la Phonation, Journal of Francais d'Oto-Rino-Laryngologie et Chirurgie Maxillo-Faciale, 9, 1960, 317-324.

12. Culver, C.A., Musical Acoustics, McGraw-Hill Book Co., New York, 1956, 21.

13. Curry, E.T., The Pitch Characteristics of the Adolescent Male Voice, Speech Monographs, 7, 1940, 48-62.

14. Curtis, J., Disorders of Voice, in Speech Handicapped School Children, Johnson, W., et al., Revised Ed., Harper Co., New York, 1956, 154.

15. Diehl, C., White, R. and Burk, K., Voice Quality and Anxiety, Journal of Speech and Hearing Research, 2, 1959, 282-285.

16. Duffy, R., The Vocal Pitch Characteristics of Eleven-, Thirteen-, and Fifteen-Year-Old Female Speakers, Unpublished Doctoral dissertation, University of Iowa, 1958.

17. Fairbanks, G., Voice and Articulation Drillbook, 2nd Ed., Harper Bros., New York, 1960, 127, 170-183.

18. Fant, G. and Risberg, A., Speech Analysis Techniques, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Interim Final Technical Report, United States Army Contract No. DA-91-591-EUC-1496, July, 1960 - June, 1961.

√ 19. Flanagan, J., Some Properties of the Glottal Sound Source, Journal of Speech and Hearing Research, 1, 1958, 99-116.

20. Fulton, R. and Trueblood, T., Practical Elements of Elocution, Boston, Ginn and Co., 1893.

21. Garrett, H.E., Statistics in Psychology and Education, 5th Ed., Longmans, Green and Co., New York, 1960, 235.

22. Goldsbury, J. and Russell, W., The American Common School Reader and Speaker, Boston, Tappan and Whittmore, 1844.

23. Gray, G. and Wise, C., The Bases of Speech, Revised Ed., Harper Bros., New York, 1946, 15.

24. Hahn, E., et al., Basic Voice Training for Speech, McGraw-Hill Book Co., New York, 1952, 67.

25. Hallock, W. and Muckey, F., The Results of 18 Years of Research Work on Voice Production and Analysis, Laryngoscope, 23, 1913, 5-20.

26. Hamill, S., The Science of Elocution, Phillips and Hunt, New York, 1882.

27. Hollien, H., An Investigation of Vocal Fry, National Institutes of Health research grant B-3396, 1963, 5-8.

28. Hollien, H., Malcik, E. and Hollien, B., Adolescent Voice Change in Southern White Males, Speech Monographs, (in press).

29. Hollien, H. and Wendahl, R., The Perception of Vocal Fry, Journal of the Acoustical Society of America, 1964, (in press).

30. Irwin, R., Speech and Hearing Therapy, Prentice-Hall, Inc., New York, 1953, 130.

31. Jackson, C. and Jackson, C., Dysphonia Plicae Ventricularis: Phonation with the Ventricular Bands, Archives of Otolaryngology, 21, 1935, 157-167.

32. Joos, M., Acoustic Phonetics, Language, 24, 1948, 6-7.

33. Lieberman, P., Perturbations in Vocal Pitch, Journal of the Acoustical Society of America, 33, 1961, 597-603.

34. Lieberman, P., Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathological Larynges, Journal of the Acoustical Soceity of America, 35, 1963, 344-353.

35. Moore, P. and Thompson, C., Comments on Physiology of Hoarseness, Archives of Otolaryngology, 1964, (in press).

36. Moore, P. and von Leden, H., Dynamic Variations of the Vibratory Pattern in the Normal Larynx, Folia Phoniatrica, 10, 1958, 205-238.

37. Moore, W., Personality Traits and Voice Quality Deficiencies, Journal of Speech Disorders, 4, 1939, 33-36.

38. Moser, H., Symposium of Unique Cases of Speech Disorders: Presentation of a Case, Journal of Speech and Hearing Disorders, 7, 1942, 173-174.

39. Mysak, E., Pitch and Duration Characteristics of Older Males, Journal of Speech and Hearing Research, 2, 1959, 46-54.

40. Peterson, G., Vowel Formant Measurement, Journal of Speech and Hearing Research, 2, 1959, 173-183.

41. Philhour, C., An Experimental Study of the Relationships between Perception of Vocal Pitch in Connected Speech and Certain Measures of Vocal Frequency, Doctoral dissertation, University of Iowa, 1948.

42. Potter, R., Kopp, G. and Green, H., Visible Speech, Van Nostrand, New York, 1947.

43. Raubicheck, L., Davis, E. and Carll, A., Voice and Speech Problems, Prentice-Hall Co., New York, 1931.

√44. Rees, M., Some Variables Affecting Perceived Harshness, Journal of Speech and Hearing Research, 1, 1958, 155-168.

45. Rush, J., The Philosophy of the Human Voice, Griggs and Elliot, 2nd Ed., Philadelphia, 1833.

46. Russell, G., Physiological Cause of Gutteral and Piercing Deaf Voices, Oralism and Auralism, 8, 1929, 98-104.

47. Russell, W. and Murdock, W., Orthophony or Vocal Culture in Elocution, Wm. D. Ticknor and Co., Boston, 1845.

48. Scripture, E., Researches in Experimental Phonetics: The Study of Speech Curves, The Carnegie Foundation, Washington, D.C., 1906.

49. Sherman, D., The Merits of Backward Playing of Connected Speech in the Scaling of Voice Quality Disorders, Journal of Speech and Hearing Disorders, 19, 1954, 312-321.

50. Sherman, D. and Jensen, P., Harshness and Oral-Reading Time, Journal of Speech and Hearing Disorders, 27, 1962, 172-177.

√51. Sherman, D. and Linke, E., The Influence of Certain Vowel Types on Degree of Harsh Voice Quality, Journal of Speech and Hearing Disorders, 17, 1952, 401-408.

52. Simon, C., The Variability of Consecutive Wave Lengths in Vocal and Instrumental Sounds, Psychological Monographs, 36, 1926, 41-83.

53. Stevens, K. and House, A., An Acoustical Theory of Vowel Production and Some of its Implications, Journal of Speech and Hearing Research, 4, 1961, 303-320.

√ 54. Thurman, W., The Construction and Acoustical Analysis of Recorded Scales of Severity for Six Voice Quality Disorders, Unpub. Doctoral dissertation, Purdue University, 1953.

55. Timke, R., von Leden, H. and Moore, P., Laryngeal Vibrations: Measurements of the Glottic Wave, Part II. Physiological Variations, Archives of Otolaryngology, 69, 1959, 438-444.

56. Van Dusen, C., A Laboratory Study of the Metallic Voice, Journal of Speech and Hearing Disorders, 6, 1941, 137-140.

57. Van Riper, C., Speech Correction: Principles and Methods, 2nd Ed., Prentice-Hall, New York, 1947, 319.

58. Van Riper, C. and Irwin, J., Voice and Articulation, Prentice-Hall, Englewood Cliffs, New Jersey, 1958, 231.

59. Voelker, C., Phoniatry in Dysphonia Ventricularis, Annals of Otology, Rhinology and Laryngology, 44, 1935, 471-473.

60. Wendahl, R., Laryngeal Analog Synthesis of Harsh Voice Quality, Folia Phoniatrica, 15, 1963, 241-250.

61. Wendahl, R., Moore, P. and Hollien, H., Comments on Vocal Fry, Folia Phoniatrica, 15, 1963, 251-255.

62. West, R., Kennedy, L. and Carr, A., The Rehabilitation of Speech, Harper Bros., New York, 1947, 137.

APPENDIX A

NAME: ____________________ DATE: ____________________

AGE: ____________________ MAJOR: ____________________

1. A B	11. A B
2. A B	12. A B
3. A B	13. A B
4. A B	14. A B
5. A B	15. A B
6. A B	16. A B
7. A B	17. A B
8. A B	18. A B
9. A B	19. A B
10. A B	20. A B

Comments:

APPENDIX B

Directions to Listeners in the Judgmental Procedure

You are going to hear a tape recording of voices from twenty different speakers. Your task is to divide these samples into two groups; there must be ten samples in each group. The basis of your division will be the quality of the voice - not the pitch nor the loudness, but the quality. Listen to the tape once and decide which voices will go into group A and which will go into group B. When the tape is played the second time, circle "A" each time you hear a voice quality you think should go into group A and circle "B" each time you hear a voice quality you think should go into group B. It does not matter which letter you assign to which group so long as you are consistent in your judgments. In other words, all the qualities in group "A" should sound either the same or similar to each other and all the qualities in group "B" should sound either the same or similar to each other. Are there any questions?

BIOGRAPHICAL SKETCH

John Frederich Michel was born June 21, 1933 at Columbus, Ohio. In June, 1951, he was graduated from South High School. From 1951 to 1955, Mr. Michel was enrolled at the Ohio State University. He served in the United States Army from January, 1955, to November, 1956, serving 17 months in Germany. Following his discharge, he completed his Bachelor of Arts degree in 1959. Mr. Michel worked for a year and a half as a research assistant in the Psycholinguistics Laboratory at the Ohio State University Research Foundation, receiving his Master of Arts degree in 1960. After two years in the graduate program at the University of Wichita, he transferred to the University of Florida where he has continued his work toward the degree of Doctor of Philosophy.

John Frederich Michel is married to the former Lorraine June Ivison. He is a member of the American Speech and Hearing Association, the Florida Speech and Hearing Association and Sigma Alpha Eta.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Arts and Sciences and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 8, 1964

E. Ruffin Jones
Dean, College of Arts and Sciences

Dean, Graduate School

Supervisory Committee:

Harry Hollien
Chairman

Richard J. Anderson

Paul D. Jensen

McKenzie Buck

Kenneth R. Bzoch

George T. Singleton